CNN-based Fraud Account Classification for Ethereum Blockchain's Accounts

Revin Naufal Alief, Syifa Maliah Rachmawati, Muhammad Rasyid Redha Ansori, Jae-Min Lee, and Dong-Seong Kim,
Networked Systems Lab., Department of IT Convergence Engineering,
Kumoh National Institute of Technology, Gumi, South Korea.
{revinnaufal, syifamr, ljmpaul, dskim}@kumoh.ac.kr

***Abstract*—Cryptocurrency is the most popular and fastest growing currency in the International Financial market today. But, as the amount of transaction is increasing, fraud accounts are also emerged that cause much loss for the ethereum account that transact each other. In the previous study, the usage of machine learning is applied to solve this problem, but the previous study only shows limited performance metrics. In this paper, a CNN-based algorithm is proposed to classify the fraud account in the ethereum network. The CNN model is applied in a dataset that contain legitimate and fraudelent transaction that is done over ethereum network. The performance shows that the CNN-based model is able to classify the fraud account successfully with an accuracy of 98.02%.**

***Index Terms*—Blockchain, Deep Learning, Ethereum, Fraud Detection**

I. Introduction

Cyrptocurrency has been attracting a lot of attention over the years. This attention is shown by the expanding number of cryptocurrencies that is emerged and also the increasing volume of transactions in the cyrptocurrency market. Customer perception of cryptocurrencies is no longer merely based on an investing excitement, but also as an evidence of stable and long-term investment [1]. The evidence of stable and long-term investment is shown by the absence of third parties, resulting the quantity supply cannot be manipulated, in contrast to fiat currencies that is vulnerable to inflation [2]. One of the most popular cryptocurrencies, Ethereum, is currently have a large volume of transaction on its network. But, unfortunately the problem still exist, especially the problem regarding the phishing scam which is considered as the biggest issue. This consideration as the biggest issue is shown by [3], that stated 50% of all cybercrimes on Ethereum since 2017 is caused by phishing scam. Thus, in order to prevent this problem, a mechanism to detect the fraud account is needed.

Currently one of the emerging technology, artificial intelligence (AI) is capable of classifying data through learning from the previous data. In blockchain application field, AI also is combined to help the improvement of the system such as unmanned aerial vehicle [4], and blockchain consensus performance [5]. It is important to be noted that AI is able to classify the fraud account based on the account's transaction history, especially using Deep Learning (DL). Compared to Machine Learning (ML), DL is considered more advanced technique of AI as it has less dependence on human interference.

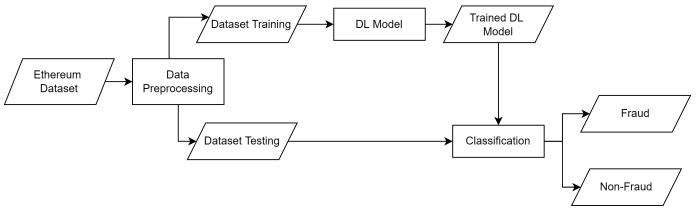

Fig. 1. Flowchart of the proposed system.

The high accuracy of fraud account classification could be acquired by creating a well design DL model. In addition, the preprocessing should be done carefully so that the DL model could learn more accurate. Before, previous studies proposed an ML-based fraud account's classification [6], [7]. But, if ML algorithm returns inaccurate prediction, human still need to intervene for fixing the problem. In contrast with DL models that allow the algorithms to determine for themselves the accuracy of predictions through their own neural networks. Thus, in the main contribution of this paper consists of:

- Proposed a DL model using CNN algorithm to classify the fraud account based on the account's transaction history.
- Comparison results with the previous existing methods.

II. Proposed System

The flowchart of proposed system in this paper is shown in Figure 3. First, the ethereum dataset is splitted into two parts for training and testing. The dataset training is used to train the DL model and after the model is trained, the model is tested to classify either fraud or non-fraud based on the testing dataset.

The dataset is obtained from [8], that contain fraud and valid transactions made over Ethereum. This dataset also used in [6] and [7] to test the machine learning model. This dataset contains 9841 transaction made over Ethereum. To be exact, this dataset consist of 7662 non-fraud transaction and 2179 fraud transaction. It has 51 features that consist of various detailed transaction, from average time took for transaction until average value of transaction. Based on these data, we do the prepocessing and use the Convolutional Neural Network model for classifying the fraud account.

The input layer configuration that is used in the model is 36 neurons. Then the layer is fitted to the Maxpooling layer

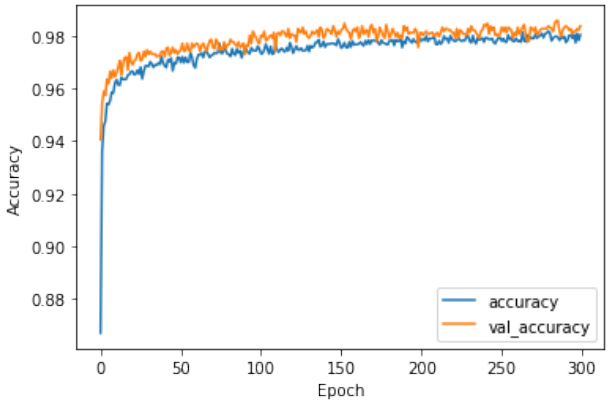

Fig. 2. The accuracy of the training and validation of the proposed model.

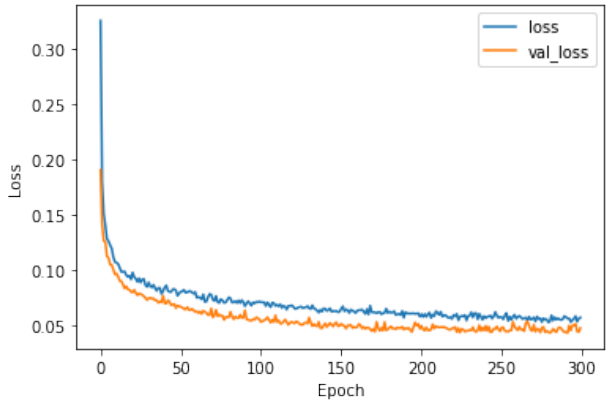

Fig. 3. The loss of training and validation loss of the proposed model.

with pool size 2. Then the result from these 2 layers are fit into the convolutional layer, to then passed through the same Maxpooling layer. Then dropout layer of 0.1 is utilized to avoid overfitting.

III. Performance Evaluation

The performance evaluation of the proposed model is shown in Figure 2 and Figure 3. Figure 2 shows the performance of the proposed model by looking at the training and validation accuracy for 300 epochs. As for Figure 3 shows the performance of the proposed model based on the loss of training and validation over 300 epochs. Based on these figures, both of the values stayed almost the same for all of the process. This shows that the CNN model able to learn the collected data properly.

TABLE I
Proposed Method Performance Evaluation

Ref	Method	Accuracy	Time (s)
[6]	XGBoost	98.15%	-
[7]	RF	97.96%	4.85
Proposed	**CNN**	**98.32%**	**0.2**

In addition, this paper tried to compare the performance in term of accuracy and time that could be seen in Table I. In term of accuracy, the proposed CNN model able to outperform the previous studies. The accuracy of proposed model shows a slight difference with [6], but in [6] the time took for classifying is not provided. Thus it is shown that the CNN model is able to outperform in term of accuracy and time compared to the previous models.

IV. Conclusion

In this paper, we tried an approach of classifying fraud account in the Ethereum blockchain framework based on the transaction history. This paper tried to apply deep learning model, to leverage machine learning model from the previous study. As deep learning models allow the algorithms to determine for themselves the accuracy of predictions through their own neural networks without human intervenes. The result shows that the DL model achieved 98.32% accuracy and 0.2 second for classifying the data. This model shows that it outperform the previous study's model. For the future works, we consider including the algorithm for preventing imbalanced dataset. As this dataset is imbalanced and the previous study has not yet considered it.

Acknowledgment

This research work was supported by Priority Research Centers Program through NRF funded by MEST (2018R1A6A1A03024003) and the Grand Information Technology Research Center support program (IITP-2022-2020-0-01612) supervised by the IITP by MSIT, Korea.

References

[1] A. Mashatan, M. S. Sangari, and M. Dehghani, "How perceptions of information privacy and security impact consumer trust in crypto-payment: An empirical study," *IEEE Access*, vol. 10, pp. 69 441–69 454, 2022.

[2] M. Hashemi Joo, Y. Nishikawa, and K. Dandapani, "Cryptocurrency, a successful application of blockchain technology," *Managerial Finance*, vol. 46, no. 6, pp. 715–733, Jan 2020. [Online]. Available: https://doi.org/10.1108/MF-09-2018-0451

[3] B. E. Mykulyak, "Facilitating online crypto-payments now and in the future," in *The PayTech Book: The Payment Technology Handbook for Investors, Entrepreneurs and FinTech Visionaries*, 2019, p. 132–133.

[4] R. Akter, M. Golam, V.-S. Doan, J.-M. Lee, and D.-S. Kim, "Iomt-net: Blockchain integrated unauthorized uav localization using lightweight convolution neural network for internet of military things," *IEEE Internet of Things Journal*, pp. 1–1, 2022.

[5] D. Kim, I. Doh, and K. Chae, "Improved raft algorithm exploiting federated learning for private blockchain performance enhancement," in *2021 International Conference on Information Networking (ICOIN)*, 2021, pp. 828–832.

[6] A. Maurya and A. Kumar, "Credit card fraud detection system using machine learning technique," in *2022 IEEE International Conference on Cybernetics and Computational Intelligence (CyberneticsCom)*, 2022, pp. 500–504.

[7] R. F. Ibrahim, A. Mohammad Elian, and M. Ababneh, "Illicit account detection in the ethereum blockchain using machine learning," in *2021 International Conference on Information Technology (ICIT)*, 2021, pp. 488–493.

[8] V. Aliyev, "Ethereum fraud detection dataset," Jan 2021. [Online]. Available: https://www.kaggle.com/datasets/vagifa/ethereum-frauddetection-dataset/